On the Whiteness of High Resolution Quantization Errors

Harish Viswanathan and Ram Zamir *
Lucent Technologies, Bell Labs, Holmdel, NJ-07733
and
Dept. of Elect. Eng. - Systems, Tel Aviv University, ISRAEL

September 11, 2000

Abstract

A common belief in quantization theory says that the quantization noise process resulting from uniform scalar quantization of a correlated discrete time process tends to be *white* in the limit of small distortion ("high resolution"). A rule of thumb for this property to hold is that the source samples have a "smooth" joint distribution. We give a precise statement of this property, and generalize it to *non-uniform* quantization and to *vector* quantization. We show that the quantization errors resulting from independent quantizations of dependent real random variables become asymptotically uncorrelated (although not necessarily statistically independent) if the joint Fisher information under translation of the two variables is finite and the quantization cells shrink uniformly as the distortion tends to zero.

Key words: high resolution, asymptotic whiteness, Fisher information, quantization noise, multiterminal source coding.

*This research was started when the authors were with Cornell University, Ithaca, NY. The second author was supported in part by the Wolfson Research Awards, administered by the Israel Academy of Science and Humanities.

I Introduction

High resolution quantization theory provides efficient design tools and explicit formulas for coding of continuous sources with small distortion [6, 5, 10]. The importance of this theory lies in the fact that similar concepts do not exist for coding at arbitrary resolution. In Shannon theory, high resolution source coding leads to useful asymptotic results for the rate-distortion function [9], and for the rate-distortion region in multiuser source coding [17, 11].

Many results and properties in high resolution quantization theory although simple to state and justify heuristically, are hard to prove rigorously. One such property is the asymptotic whiteness of the quantization error process [6, sec. 5.6], which says, in effect, that independent "fine" quantization of two "smoothly" dependent random variables generates *un*correlated errors. The Asymptotic Whiteness Property (AWP) plays an important role in practical speech and image compression, where the quantization noise spectrum has a strong perceptual effect [8]. The AWP also gives interesting insight into the behavior of multiterminal coding of correlated continuous sources [1, 16, 17], where the correlation between the errors at separate terminals may affect the estimation error at the centralized decoder.

Consider, for example, scalar quantization of a stationary process $X = X_1, X_2, \ldots$ using a uniform quantizer $Q_u(\cdot)$ with step size Δ, i.e.,

$$Q_u(x_n) = \arg \min_{\hat{x} \in \{0, \pm\Delta, \pm 2\Delta, \ldots\}} |x_n - \hat{x}| \qquad n = 1, 2, \ldots . \tag{1}$$

Let $Z_n = Q_u(X_n) - X_n$ denote the error in quantizing the n-th sample of X. The AWP implies that under certain conditions on the pairwise distribution of X, the correlation coefficient between any two quantization error samples satisfies

$$\rho_k \stackrel{\Delta}{=} \frac{EZ_nZ_{n+k}}{EZ_n^2} \to 0 \quad \text{as } \Delta \to 0 \text{ for all } k \neq 0. \tag{2}$$

Under these conditions the mean squared error in quantization satisfies $EZ_n^2 = D \approx \Delta^2/12$ [12], where throughout the paper $\approx$ means that the ratio between the corresponding quantities goes to 1, thus (2) amounts to $EZ_nZ_{n+k}/\Delta^2 \to 0$ as $\Delta \to 0$.

The uniform scalar quantizer above provides a simple example that captures the spirit of the AWP. The AWP extends straightforwardly with simple conditions to the case of vector lattice quantization that subsumes the scalar quantization example above. Specifically, we show rigorously in Theorem 1 that the correlation coefficient between lattice quantization error vectors vanishes asymptotically under the simple condition that source density exists. The regularity of the quantization scheme combined with this condition is sufficient to guarantee the AWP. Related work

regarding second moments of lattice quantization noise appears in [10, 12, 18].

Our main result in this paper (Theorem 2) generalizes the AWP to *non-uniform* quantization. Unlike for the lattice quantization case, in this case the quantization cells are not necessarily convex, and may be even unions of disconnected regions, as happens in the case of multiterminal source coding [17]. This more general formulation of the AWP requires stronger conditions on the joint distribution of (X_n, X_{n+k}).

Non-uniform quantization provides a particularly interesting implication of the AWP in which correlation and statistical dependence play different roles. It is not hard to construct examples for a source and a non-uniform quantizer such that the quantization errors would be strongly statistically dependent, but still asymptotically uncorrelated.

The intuition behind the AWP comes from the combination of two ideas:

1. **Local uniformity:** If the joint distribution of the source samples is "smooth", then it is approximately uniform inside small cells (corresponding to high resolution quantization).

2. **Rectangular partition:** Independent quantization of random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ induces a rectangular ("Cartesian") partitioning of the $(\mathcal{X}, \mathcal{Y})$-plane. If furthermore, the joint distribution of (X, Y) is *exactly* uniform in some rectangular cell, then the quantization errors are conditionally independent given that cell. If the joint distribution is piecewise uniform with respect to all cells, then the overall quantization errors are un-correlated (although not necessarily statistically independent).

The property of rectangular partition above seems simple and clear. The main purpose of this paper is to make a precise statement of the idea of local uniformity, to propose a sufficient condition for it to hold and to prove a general form of the AWP using the local uniformity condition. For lattice quantization existence of the joint probability density of the source turns out to be sufficient, as stated in Theorem 1. For general non-uniform quantization our condition is based on the finiteness of the *Fisher Information under translation*, a quantity which is a function of the joint distribution of the source samples. Lemma 1 provides an intermediate result on local uniformity in terms of the Fisher Information, while Theorem 2 states the AWP for non-uniform quantizers. Section II summarizes these results and Section III provides the proofs.

II Summary of Results

A. Lattice Quantization

A *lattice* quantizer is a vector quantizer in which the quantization cells are Voronoi regions of a $k-$dimensional lattice and the reconstructions $\hat{\mathbf{x}}$ are geometric centroids of the cells [12]. If the polytope P is the Voronoi region of the lattice, then each cell in the quantizer is congruent to P. A sequence of lattice quantization schemes $Q^x_{\alpha_N}$ with distortion tending to zero is obtained with quantization regions congruent to $P_{\alpha_N} = \{\alpha_N \mathbf{x} : \mathbf{x} \in P\}$, $\alpha_N > 0$ and $\alpha_N \to 0$ as $N \to \infty$.

Consider a sequence of pairs of lattice quantizers $Q^x_{\alpha_N}$ and $Q^y_{\alpha_N}$, not necessarily derived from the same lattice, with reconstructions $\hat{\mathbf{x}}_N$ and $\hat{\mathbf{y}}_N$, respectively. Denote the distortion of such a sequence by

$$D_{x,N} \triangleq E\|\mathbf{X} - \hat{\mathbf{X}}_N\|^2, \quad D_{y,N} \triangleq E\|\mathbf{Y} - \hat{\mathbf{Y}}_N\|^2. \tag{3}$$

Let

$$\rho_N \triangleq \frac{E\{(\mathbf{X} - \hat{\mathbf{X}}_N)^t(\mathbf{Y} - \hat{\mathbf{Y}}_N)\}}{\sqrt{D_{x,N} D_{y,N}}} \tag{4}$$

be the correlation coefficient between the quantization errors, where $(\cdot)^t$ denotes transpose. We have the following result on the correlation coefficient between the $\mathbf{X}, \mathbf{Y}$ quantization errors.

Theorem 1 (AWP for Lattice Quantizers) *Let $(\mathbf{X}, \mathbf{Y}) \in (\mathcal{X}, \mathcal{Y})$, where $\mathcal{X} = \mathcal{Y} = \mathcal{R}^k$, be correlated random vectors with source density $p(\mathbf{x}, \mathbf{y})$. For the sequence of pairs of lattice quantizers with distortions tending to zero,*

$$\rho_N \to 0 \quad \textit{as} \quad N \to \infty.$$

The basic form of the AWP in (2), regarding uniform quantization of a stationary source with a density, follows as a corollary from this theorem (setting $k = 1$, $X = X_n$ and $Y = X_{n+k}$). Note that Theorem 1 can also be deduced from the analysis in [10] on multidimensional companding with locally quadratic distortion measures (see also [12]).

B. Non-uniform Quantization

We first define the Fisher Information under translation of $\mathbf{X}$ (or in short, the FI of $\mathbf{X}$). Let $\mathbf{X} \in \mathcal{R}^m$ have a density $p(\mathbf{x})$, and define the FI of $\mathbf{X}$ as [2, 4, 3]

$$J(\mathbf{X}) \triangleq \int \frac{1}{p(\mathbf{x})} \cdot \left\| \frac{\partial p(\mathbf{x})}{\partial \mathbf{x}} \right\|^2 d\mathbf{x} \triangleq E\left\{ \left\| \frac{\partial \ln p(\mathbf{X})}{\partial \mathbf{X}} \right\|^2 \right\} \tag{5}$$

where $\frac{\partial p(\mathbf{x})}{\partial \mathbf{x}}$ (respct. $\frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}}$) denotes the gradient (i.e., vector of partial derivatives) of $p(\mathbf{x})$ (respct. of $\ln p(\mathbf{x})$) with respect to $x_1, \ldots, x_m$, $\|\cdot\|$ denotes the Euclidean norm, and $E\{\}$ denotes expectation

with respect to $\mathbf{X}$. Finite FI means that the quantity $\left\| \frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}} \right\| = \frac{1}{p(\mathbf{x})} \cdot \left\| \frac{\partial p(\mathbf{x})}{\partial \mathbf{x}} \right\|$ is finite for most probable $\mathbf{x}$s. Roughly, this implies that the *relative variation* of the density

$$\max_{\mathbf{x},\mathbf{y} \in T} \frac{p(\mathbf{x}) - p(\mathbf{y})}{p(\mathbf{y})}$$

inside a small cell $T \subset \mathcal{R}^m$ is small, i.e., $p(\mathbf{x})$ is locally uniform inside T.

The following lemma suggests a way to quantify the degree of local uniformity of the source, relative to a "fine" vector quantizer (not necessarily a lattice), in terms of the FI of the source and the quantizer's distortion:

Lemma 1 (Local Uniformity) *Let $Q : \mathcal{R}^m \to \mathcal{R}^m$ be a vector quantizer which encodes $\mathbf{X} \in \mathcal{R}^m$ with mean squared error distortion D, i.e., $E\|Q(\mathbf{X}) - \mathbf{X}\|^2 = D$. Let $p(\mathbf{x})$ denote the density of $\mathbf{X}$. If the norm of the Hessian of $\ln p(\mathbf{x})$ is bounded, i.e., $\left\| \frac{\partial^2 \ln p(\mathbf{x})}{\partial \mathbf{x}^2} \right\| < K$ for all $\mathbf{x}$ in $\mathcal{R}^m$ for some K, then as $D \to 0$*

$$E\left\{ \left| \ln \left(\frac{p(\mathbf{X})}{p(Q(\mathbf{X}))} \right) \right| \right\} \leq \sqrt{J(\mathbf{X}) \cdot D} + O(D), \tag{6}$$

where $O(D) = \frac{K}{2} D$, and $J(\mathbf{X})$ is the FI of the source.

Note that for small D the term $\sqrt{J(\mathbf{X})D}$ dominates the bound. Thus, if the FI is finite and the distortion is small, then the ratio of the density at an arbitrary point $\mathbf{x}$ to the density at the reconstruction point $Q(\mathbf{x})$ is on the average close to one, or equivalently "$p(\mathbf{x})$ is locally uniform".

As discussed in the Introduction, our main result uses the finiteness of the joint FI to show that errors resulting from independent quantization of source samples are asymptotically uncorrelated. The following setting generalizes the lattice quantization scheme above to non-uniform vector quantization with joint centroid decoding.

Let $\mathbf{X} \in \mathcal{X}$, $\mathbf{Y} \in \mathcal{Y}$, where $\mathcal{X} = \mathcal{Y} = \mathcal{R}^k$, be random vectors with joint density $p(\mathbf{x}, \mathbf{y})$. Let

$$i(\mathbf{x}) : \mathcal{X} \to \{1, 2, \ldots, N_x\},$$

$$j(\mathbf{y}) : \mathcal{Y} \to \{1, 2, \ldots, N_y\} \tag{7}$$

induce two partitions of $\mathcal{R}^k$ corresponding to *independent* quantization of $\mathbf{X}$ and $\mathbf{Y}$, respectively. We do not exclude the case where N_x or N_y are infinite. Let

$$T_i^x = \{\mathbf{x} : i(\mathbf{x}) = i\} \quad \text{and} \quad T_j^y = \{\mathbf{y} : j(\mathbf{y}) = j\}$$

denote the cells of these partitions, and let

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = Q\Big(i(\mathbf{x}), j(\mathbf{y})\Big), \tag{8}$$

where $Q:\ \{1,2,\ldots,N_x\}\times\{1,2,\ldots,N_y\}\to(\hat{\mathcal{X}},\hat{\mathcal{Y}})$ denotes the quantizer reconstruction function. We define $Q(i,j)$ as the joint centroid of the cell $T_i^x\times T_j^y$ relative to the source distribution, i.e.,

$$i(\mathbf{x})=i, j(\mathbf{y})=j \quad \Longrightarrow \quad \begin{array}{l}\hat{\mathbf{x}}=E\Big\{\mathbf{X}|\ (\mathbf{X},\mathbf{Y})\in T_i^x\times T_j^y\Big\}\\ \hat{\mathbf{y}}=E\Big\{\mathbf{Y}|\ (\mathbf{X},\mathbf{Y})\in T_i^x\times T_j^y\Big\}\end{array} \tag{9}$$

where here $E\{\cdot\}$ denotes conditional expectation with respect to the *true* distribution of $(\mathbf{X},\mathbf{Y})$ inside the cell $T_i^x\times T_j^y$.

Note that the encoding of $\mathbf{x}$ and the encoding of $\mathbf{y}$ as in (7) are independent; however, as opposed to the case of lattice quantization, the reconstruction function (9) depends on *both* marginal partition indices, and depends on the joint distribution of $(\mathbf{X},\mathbf{Y})$. This reconstruction function is typical of distribution optimized quantization [8], or of joint decoding in multiterminal source coding [1]. This reconstruction function coincides with the geometric centroid (as in the lattice case) for a cell $T_i^x\times T_j^y$, if the joint conditional distribution of $(\mathbf{X},\mathbf{Y})$ inside this cell is exactly uniform.

Consider a sequence of pairs of partition functions $i_N(\mathbf{x}), j_N(\mathbf{y})$ of $\mathcal{X},\mathcal{Y}$, $N=1,2,\ldots$, and a corresponding sequence of reconstruction functions $(\hat{\mathbf{x}}_N,\hat{\mathbf{y}}_N)$, such that

$$D_{x,N}\triangleq E\|\mathbf{X}-\hat{\mathbf{X}}_N\|^2\to 0, \quad D_{y,N}\triangleq E\|\mathbf{Y}-\hat{\mathbf{Y}}_N\|^2\to 0 \tag{10}$$

at the same rate. We make the following assumption on the asymptotic distribution of the quantization errors as $N\to\infty$.

Moment Condition: There exists some $\delta>0$ such that

$$\limsup_{N\to\infty}E\left(\frac{\|\mathbf{X}-\hat{\mathbf{X}}_N\|^2}{D_{x,N}}\right)^{1+\delta}<\infty, \quad \limsup_{N\to\infty}E\left(\frac{\|\mathbf{Y}-\hat{\mathbf{Y}}_N\|^2}{D_{y,N}}\right)^{1+\delta}<\infty. \tag{11}$$

The moment condition guarantees that the cells shrink uniformly over a large enough region of the support of the density $p(\mathbf{x},\mathbf{y})$. For example, the moment condition holds for uniform / lattice quantization provided the source has a density. Some easy to check conditions that imply the moment condition are given in the Appendix. We emphasize that for general partitions the moment condition is needed even when the source density has a bounded support. A similar assumption (UACI-p hypothesis) on the quantization error has been made in [13].

Our main result in the paper is summarized in Theorem 2 below. It states that under the assumptions above, the correlation coefficient between the $\mathbf{X},\mathbf{Y}$ quantization errors vanishes asymptotically, provided that the joint FI of $(\mathbf{X},\mathbf{Y})$ is finite.

Theorem 2 (AWP for Non-Uniform Quantizers) *Let $(\mathbf{X}, \mathbf{Y}) \in (\mathcal{X}, \mathcal{Y})$, where $\mathcal{X} = \mathcal{Y} = \mathcal{R}^k$, be correlated random vectors with continuous source density $p(\mathbf{x}, \mathbf{y})$, a.s. continuously differentiable $\ln p(\mathbf{x}, \mathbf{y})$ and joint FI $J(\mathbf{X}, \mathbf{Y}) < \infty$ as defined in (5). Let $i_N(\mathbf{x})$ and $j_N(\mathbf{y})$ be a sequence of independent partition functions of $\mathcal{X}$ and $\mathcal{Y}$, let $\hat{\mathbf{x}}_N$ and $\hat{\mathbf{y}}_N$ be the corresponding reconstructions (9), and let ρ_N be the correlation coefficient between the quantization errors as in (4). If the sequence $(\hat{\mathbf{x}}_N, \hat{\mathbf{y}}_N)$ satisfies (10) and (11), then*

$$\rho_N \to 0 \quad as \quad N \to \infty. \tag{12}$$

III Derivation of Results for Lattice Quantization

Proof of Theorem 1

For lattice quantization we have,

$$\|\mathbf{x} - \hat{\mathbf{x}}_N\| \leq \alpha_N \mathrm{diam}(P)$$

where P is the polytope associated with the Voronoi region of the lattice. Furthermore, Linder and Zeger [12] have shown that for tessellating quantizers and in particular, for $k-$dimensional lattice quantizers

$$\lim_{\alpha_N \to 0} \frac{D_{x,N}}{\alpha_N^2} = \frac{1}{k\mathcal{V}(P)} \int_P \|\mathbf{x} - \hat{\mathbf{x}}\|^2 d\mathbf{x}$$

where $\mathcal{V}(P)$ is the volume of P and $\hat{x}$ is the centroid of P. Hence it follows that

$$\begin{aligned} \frac{\|\mathbf{x} - \hat{\mathbf{x}}_N\|}{\sqrt{D_{x,N}}} &\leq \frac{\alpha_N \mathrm{diam}(P)}{\sqrt{D_{x,N}}} \\ &\leq 2\sqrt{\frac{k\mathcal{V}(P)}{\int_P \|\mathbf{x} - \hat{\mathbf{x}}\|^2 d\mathbf{x}}} \mathrm{diam}(P) \end{aligned}$$

for sufficiently large N. From a similar statement for $\mathbf{Y}$, we thus have for some M, $\frac{\|\mathbf{x}-\hat{\mathbf{x}}_N\|}{\sqrt{D_{x,N}}} \leq M$ and $\frac{\|\mathbf{y}-\hat{\mathbf{y}}_N\|}{\sqrt{D_{y,N}}} \leq M$ for all sufficiently large N.

Let, for $(\mathbf{x}, \mathbf{y}) \in T_i^x \times T_j^y$,

$$\begin{aligned} \tilde{p}^N(\mathbf{x}, \mathbf{y}) &\triangleq \frac{\int_{T_i^x \times T_j^y} p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y}}{\int_{T_i^x \times T_j^y} d\mathbf{x} d\mathbf{y}} \\ &\triangleq p_{i,j}. \end{aligned}$$

Then,

$$\begin{aligned} |\rho_N| &= \left| \int \frac{(\mathbf{x} - \hat{\mathbf{x}}_N)^t (\mathbf{y} - \hat{\mathbf{y}}_N)}{\sqrt{D_{x,N} D_{y,N}}} p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \right| \\ &\leq M^2 \int \left| p(\mathbf{x}, \mathbf{y}) - \tilde{p}^N(\mathbf{x}, \mathbf{y}) \right| d\mathbf{x} d\mathbf{y} \end{aligned}$$

$$
\begin{aligned}
&+ \left| \sum_{i,j} p_{i,j} \int_{\mathbf{x}\in T_i^x} \frac{(\mathbf{x}-\hat{\mathbf{x}}_N)^t}{\sqrt{D_{x,N}}} d\mathbf{x} \int_{\mathbf{y}\in T_i^y} \frac{(\mathbf{y}-\hat{\mathbf{y}}_N)^t}{\sqrt{D_{y,N}}} \right| d\mathbf{y} \\
&= M^2 \int \left| p(\mathbf{x},\mathbf{y}) - \tilde{p}^N(\mathbf{x},\mathbf{y}) \right| d\mathbf{x} d\mathbf{y} \\
&\to 0 \quad \text{as } N \to \infty
\end{aligned}
$$

where the first term vanishes since each of the integrals is zero for every (i,j) and the second term vanishes since by Lebesgue's differentiation theorem [15], $\tilde{p}^N \to p$ as $\alpha_N \to 0$, which via Scheffe's theorem [7] implies convergence in mean. □

IV Derivation of Results for Non-uniform Quantization

Proof of Lemma 1

By Taylor expansion in the Lagrange remainder form ([14]) we have

$$
\ln\left(\frac{p(\mathbf{x})}{p(Q(\mathbf{x}))}\right) = (\mathbf{x}-Q(\mathbf{x}))^t \left(\frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}}\right) + \frac{1}{2}(\mathbf{x}-Q(\mathbf{x}))^t \frac{\partial^2 \ln p(\mathbf{z})}{\partial \mathbf{z}^2}(\mathbf{x}-Q(\mathbf{x}))
$$

for some vector $\mathbf{z}$ between $\mathbf{x}$ and $Q(\mathbf{x})$. Now,

$$
\begin{aligned}
(\mathbf{x}-Q(\mathbf{x}))^t \frac{\partial^2 \ln p(\mathbf{z})}{\partial \mathbf{z}^2}(\mathbf{x}-Q(\mathbf{x})) &\leq \|(\mathbf{x}-Q(\mathbf{x}))\| \cdot \|\frac{\partial^2 \ln p(\mathbf{z})}{\partial \mathbf{z}^2}\| \cdot \|(\mathbf{x}-Q(\mathbf{x}))\| \\
&\leq K\|\mathbf{x}-Q(\mathbf{x})\|^2
\end{aligned}
$$

where the last inequality is from the boundedness assumption of the Lemma. Hence, applying the Cauchy Schwarz inequality to the first term we get

$$
\begin{aligned}
\int \left| \ln \frac{p(\mathbf{x})}{p(Q(\mathbf{x}))} \right| p(\mathbf{x}) d\mathbf{x} &\leq \int \left| (\mathbf{x}-Q(\mathbf{x}))^t (\frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}}) \right| p(\mathbf{x}) d\mathbf{x} + \frac{1}{2}\int K||\mathbf{x}-Q(\mathbf{x})||^2 p(\mathbf{x}) d\mathbf{x} \\
&\leq \left(\int ||\mathbf{x}-Q(\mathbf{x})||^2 p(\mathbf{x}) d\mathbf{x} \right)^{1/2} \left(\int \|\frac{\partial \ln p(\mathbf{x})}{\partial x}\|^2 p(\mathbf{x}) d\mathbf{x} \right)^{1/2} + \\
&\quad + \frac{K}{2} \int ||\mathbf{x}-Q(\mathbf{x})||^2 p(\mathbf{x}) d\mathbf{x} \\
&= \sqrt{DJ(\mathbf{X})} + \frac{K}{2} D
\end{aligned}
$$

where the last equality follows from the definition of $J(\mathbf{X})$. □

Proof of Theorem 2

Fix N, and let $\mathcal{A}_N$ be a finite subset of $\mathcal{R}^k \times \mathcal{R}^k$ to be specified later. Define $\tilde{p}^N(\mathbf{x},\mathbf{y})$ to be a piecewise uniform approximation of $p(\mathbf{x},\mathbf{y})$ inside $\mathcal{A}_N$, relative to the partition function $i_N(\mathbf{x}) \times$

$j_N(\mathbf{y})$, and a copy of $p(\mathbf{x},\mathbf{y})$ outside $\mathcal{A}_N$, i.e.,

$$\tilde{p}^N(\mathbf{x},\mathbf{y}) \triangleq \begin{cases} \frac{\int_{\mathcal{A}_N \cap T_i^x \times T_j^y} p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y}}{\int_{\mathcal{A}_N \cap T_i^x \times T_j^y} d\mathbf{x}d\mathbf{y}} = \frac{P_{\mathbf{X},\mathbf{Y}}[\mathcal{A}_N \cap T_i^x \times T_j^y]}{\mathcal{V}(\mathcal{A}_N \cap T_i^x \times T_j^y)} & \text{for } (\mathbf{x},\mathbf{y}) \in \mathcal{A}_N \cap T_i^x \times T_j^y \\ p(\mathbf{x},\mathbf{y}) & \text{for } (\mathbf{x},\mathbf{y}) \in \mathcal{A}_N^c \end{cases} \tag{13}$$

where $P_{\mathbf{X},\mathbf{Y}}[T]$ denotes the probability of the set T induced by $(\mathbf{X},\mathbf{Y})$, and $\mathcal{V}(\cdot)$ denotes volume.

Define the pointwise correlation coefficient

$$\rho_N(\mathbf{x},\mathbf{y}) \triangleq \frac{(\mathbf{x}-\hat{\mathbf{x}}_N)^t(\mathbf{y}-\hat{\mathbf{y}}_N)}{\sqrt{D_{x,N}D_{y,N}}}.$$

Starting from the definition of ρ_N we have,

$$\begin{aligned} \rho_N &= \int \rho_N(\mathbf{x},\mathbf{y}) p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y} \\ &= E_{\tilde{p}^N}[\rho_N(\mathbf{X},\mathbf{Y})] + \int_{\mathcal{A}_N} \rho_N(\mathbf{x},\mathbf{y}) \left(p(\mathbf{x},\mathbf{y}) - \tilde{p}^N(\mathbf{x},\mathbf{y})\right) d\mathbf{x}d\mathbf{y} \end{aligned}$$

where $E_{\tilde{p}^N}[\cdot]$ denotes expectation w.r.t the distribution $\tilde{p}^N$ defined in (13).

Let $M_N \in \mathcal{R}$ be a sequence to be specified later. We have

$$\begin{aligned} \rho_N &= E_{\tilde{p}^N}[\rho_N(\mathbf{X},\mathbf{Y})] + \int_{\{|\rho_N(\mathbf{x},\mathbf{y})| \le M_N\} \cap \mathcal{A}_N} \rho_N(\mathbf{x},\mathbf{y}) \left(p(\mathbf{x},\mathbf{y}) - \tilde{p}^N(\mathbf{x},\mathbf{y})\right) d\mathbf{x}d\mathbf{y} \\ &\quad + \int_{\{|\rho_N(\mathbf{x},\mathbf{y})| > M_N\} \cap \mathcal{A}_N} \rho_N(\mathbf{x},\mathbf{y}) \left(p(\mathbf{x},\mathbf{y}) - \tilde{p}^N(\mathbf{x},\mathbf{y})\right) d\mathbf{x}d\mathbf{y} \end{aligned}$$

whose absolute value can be bounded as

$$|\rho_N| \le \left|E_{\tilde{p}^N}[\rho_N(\mathbf{X},\mathbf{Y})]\right| \tag{14a}$$

$$+ M_N \int_{\mathcal{A}_N} \left|p(\mathbf{x},\mathbf{y}) - \tilde{p}^N(\mathbf{x},\mathbf{y})\right| d\mathbf{x}d\mathbf{y} \tag{14b}$$

$$+ \int_{\{|\rho_N(\mathbf{x},\mathbf{y})| > M_N\} \cap \mathcal{A}_N} |\rho_N(\mathbf{x},\mathbf{y})| p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y} \tag{14c}$$

$$+ \int_{\{|\rho_N(\mathbf{x},\mathbf{y})| > M_N\} \cap \mathcal{A}_N} |\rho_N(\mathbf{x},\mathbf{y})| \tilde{p}^N(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y} \ . \tag{14d}$$

The motivation to break up the cross-correlation into four parts as in (14) is that it would allow us to apply the notions of "local uniformity", "rectangular partition", and the "moment condition" to show that the cross-correlation tends to zero at high resolution. Roughly, (14a) will tend to zero at high resolution due to $\tilde{p}^N$ being uniform in rectangular cells over most part of $\mathcal{R}^{2k}$. For appropriately chosen sequence of sets $\mathcal{A}_N$ and constants M_N (see (17) and Proposition 1), (14b) will tend to zero as $\tilde{p}^N$ better approximates p at high resolution conditions. Finally, the terms (14c) and (14d) vanish due to the "moment condition". We now present a rigorous analysis of the various terms. The following propositions lead to the fact that each of the four terms in (14) tends to zero as $D_{x,N}, D_{y,N} \to 0$.

Define $r_x(\mathcal{A}_N)$ and $r_y(\mathcal{A}_N)$ to be the maximum radii of the $\mathbf{x}-$cells and the $\mathbf{y}-$cells inside $\mathcal{A}_N$, respectively:

$$\begin{aligned} r_x(\mathcal{A}_N) &\stackrel{\Delta}{=} \sup\{\|\mathbf{x}-\hat{\mathbf{x}}_N\| : (\mathbf{x},\mathbf{y}) \in \mathcal{A}_N\} \\ r_y(\mathcal{A}_N) &\stackrel{\Delta}{=} \sup\{\|\mathbf{y}-\hat{\mathbf{y}}_N\| : (\mathbf{x},\mathbf{y}) \in \mathcal{A}_N\} \end{aligned}$$

where $(\hat{\mathbf{x}}_N, \hat{\mathbf{y}}_N) = Q(i_N(\mathbf{x}), j_N(\mathbf{y}))$. We are now ready to state and prove the prove the propositions leading to the proof of Theorem 2.

Proposition 1 (Choice of "good" $\mathcal{A}_N$) *Under the moment condition, for a continuous source density $p(\mathbf{x},\mathbf{y})$ and continuously differentiable $\ln p(\mathbf{x},\mathbf{y})$ with $J(\mathbf{X},\mathbf{Y}) < \infty$ there exists a sequence of sets $\mathcal{A}_N$ and $\epsilon > 0$ such that, as $N \to \infty$,*

1. $P(\mathcal{A}_N^c) = O(D_{x,N}^{\epsilon(1+\delta)} + D_{y,N}^{\epsilon(1+\delta)})$
2. $r_x^2(\mathcal{A}_N) = O(D_{x,N}^{1-\epsilon})$
3. $r_y^2(\mathcal{A}_N) = O(D_{y,N}^{1-\epsilon})$
4. $\sup_{\mathcal{A}_N} \|\frac{\partial \ln p(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\| = O(D_{x,N}^{-\epsilon(1+\delta)} D_{y,N}^{-\epsilon(1+\delta)})$.

Furthermore, $\mathcal{A}_N$ can be chosen so that $\mathcal{A}_N \cap T_i^x \times T_j^y$ is a rectangle $S_i^x \times S_j^y$.

Proof: Let $r_{x,N}, r_{y,N}, t_N$ be three sequences of real numbers such that $r_{x,N}^2 = \Theta(D_{x,N}^{1-\epsilon}), r_{y,N}^2 = \Theta(D_{y,N}^{1-\epsilon})$ and, $t_N = \Theta(D_{x,N}^{-\epsilon(1+\delta)} D_{y,N}^{-\epsilon(1+\delta)})$ where the notation $x_N = \Theta(y_N)$ is used to imply that $0 < \liminf \frac{x_N}{y_N} \leq \limsup \frac{x_N}{y_N} < \infty$. Let

$$\begin{aligned} \mathcal{A}_N^1 &= \{(\mathbf{x},\mathbf{y}) : \|\mathbf{x}-\hat{\mathbf{x}}_N\| \leq r_{x,N}\} \\ \mathcal{A}_N^2 &= \{(\mathbf{x},\mathbf{y}) : \|\mathbf{y}-\hat{\mathbf{y}}_N\| \leq r_{y,N}\} \\ \mathcal{A}_N^3 &= \{(\mathbf{x},\mathbf{y}) : \|\frac{\partial \ln p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\| \leq t_N\} \end{aligned}$$

From the definition of $\mathcal{A}_1$ and the moment condition we have

$$\begin{aligned} r_{x,N}^{2+2\delta} P((\mathcal{A}_N^1)^c) &\leq E[\|\mathbf{X}-\hat{\mathbf{X}}_N\|^{2+2\delta}] \\ &= O(D_{x,N}^{1+\delta}). \end{aligned}$$

Hence

$$\begin{aligned} P((\mathcal{A}_N^1)^c) &\leq \frac{O(D_{x,N}^{1+\delta})}{D_{x,N}^{(1-\epsilon)(1+\delta)}} \\ &= O(D_{x,N}^{\epsilon(1+\delta)}). \end{aligned}$$

Similarly,

$$P((\mathcal{A}_N^2)^c) \leq O(D_{y,N}^{\epsilon(1+\delta)}).$$

Note that

$$\begin{aligned} E\left[\|\frac{\partial \ln p_{X,Y}(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\|\right] &\leq \left(E\left[\|\frac{\partial \ln p_{X,Y}(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\|^2\right]\right)^{1/2} \\ &= \sqrt{J(\mathbf{X},\mathbf{Y})} < \infty \end{aligned}$$

and hence

$$P_{X,Y}[\{(\mathbf{x},\mathbf{y}) : \|\frac{\partial \ln p(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\| > t_N\}] \leq \frac{\sqrt{J(\mathbf{X},\mathbf{Y})}}{t_N} \tag{15}$$

or

$$P[\|\frac{\partial \ln p(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\| > t_N] \leq O(D_{x,N}^{\epsilon(1+\delta)} D_{y,N}^{\epsilon(1+\delta)})$$

Choose $\mathcal{A}_N = \mathcal{A}_N^1 \cap \mathcal{A}_N^2 \cap \mathcal{A}_N^3$. It follows that

$$\begin{aligned} P(\mathcal{A}_N^c) &\leq P((\mathcal{A}_N^1)^c) + P((\mathcal{A}_N^2)^c) + P((\mathcal{A}_N^3)^c) \\ &= O(D_{x,N}^{\epsilon(1+\delta)} + D_{y,N}^{\epsilon(1+\delta)}), \\ r_x(\mathcal{A}_N)^2 &\leq r_{x,N}^2 \\ &= O(D_x^{1-\epsilon}), \\ r_y(\mathcal{A}_N)^2 &\leq r_{y,N}^2 \\ &= O(D_y^{1-\epsilon}), \\ \sup_{\mathcal{A}_N} \|\frac{\partial \ln p(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\| &\leq t_N \\ &= O(D_{x,N}^{-\epsilon(1+\delta)} D_{y,N}^{-\epsilon(1+\delta)}). \end{aligned}$$

Thus we have demonstrated by construction that it is possible to find a sequence of $\mathcal{A}_N$s simultaneously satisfying conditions 1, 2, 3 and 4 in the statement of the proposition.

Suppose $\mathcal{A}_N \cap T_i^x \times T_j^y$ is not a rectangle for some cell (i,j). Then for some $(\mathbf{x}_o,\mathbf{y}_o) \in \mathcal{A}_N \cap T_i^x \times T_j^y$

$$\mathcal{A}_N^{'} = \mathcal{A}_N \cup (\{(\mathbf{x},\mathbf{y}) : \|\mathbf{x}-\mathbf{x}_o\| < 2r_{x,N}, \|\mathbf{y}-\mathbf{y}_o\| < 2r_{y,N}\} \cap (T_i^x \times T_j^y))$$

guarantees $\mathcal{A}_N^{'} \cap T_i^x \times T_j^y$ is a rectangle without violating conditions 1, 2, 3, and furthermore, condition 4 continues to hold because of the continuity of $\frac{\partial \ln p(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}$. □

Proposition 2 (Regarding (14b))

$$\int |p(\mathbf{x},\mathbf{y}) - \tilde{p}^N(\mathbf{x},\mathbf{y})| d\mathbf{x} d\mathbf{y} \leq \sqrt{4\sqrt{r_x(\mathcal{A}_N)^2 + r_y(\mathcal{A}_N)^2}\left(\sup_{\mathcal{A}_N} \|\frac{\partial \ln p(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\|\right)}.$$

Proof : Since $p(\mathbf{x},\mathbf{y})$ is a continuous function, by the Mean value theorem ([14]) there exists a point $(\mathbf{x}_{ij},\mathbf{y}_{ij})$ in the convex hull of $\mathcal{A}_N \cap (T_i^x \times T_j^y)$ such that

$$\begin{aligned}\tilde{p}^N(\mathbf{x},\mathbf{y}) &\triangleq \frac{\int_{\mathcal{A}_N \cap T_i^x \times T_j^y} p(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}}{\mathcal{V}(\mathcal{A}_N \cap T_i^x \times T_j^y)} \\ &= p(\mathbf{x}_{ij},\mathbf{y}_{ij}) \quad \text{for all } (\mathbf{x},\mathbf{y}) \in \mathcal{A}_N \cap (T_i^x \times T_j^y)\end{aligned}$$

Hence the divergence (in nats) between p and $\tilde{p}^N$ satisfies, [3],

$$\begin{aligned}D\left(p\|\tilde{p}^N\right) &= \int_{\mathcal{A}_N} p(\mathbf{x},\mathbf{y}) \ln \frac{p(\mathbf{x},\mathbf{y})}{\tilde{p}^N(\mathbf{x},\mathbf{y})} d\mathbf{x}d\mathbf{y} \\ &\leq \int_{\mathcal{A}_N} p(\mathbf{x},\mathbf{y}) \left|\ln \frac{p(\mathbf{x},\mathbf{y})}{\tilde{p}^N(\mathbf{x},\mathbf{y})}\right| d\mathbf{x}d\mathbf{y} \\ &= \sum_{i,j} \int_{\mathcal{A}_N \cap T_i^x \times T_j^y} p(\mathbf{x},\mathbf{y}) \left|\ln \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x}_{ij},\mathbf{y}_{ij})}\right| d\mathbf{x}d\mathbf{y}.\end{aligned}$$

Now by Taylor Series expansion in Lagrange remainder form (see Rudin [1]) we have

$$\begin{aligned}&|\ln p(\mathbf{x},\mathbf{y}) - \ln p(\mathbf{x}_{ij},\mathbf{y}_{ij})| = \\ &= (\mathbf{x}-\mathbf{x}_{ij})^t \left(\frac{\partial \ln p(\tilde{\mathbf{x}}_{ij},\tilde{\mathbf{y}}_{ij})}{\partial \mathbf{x}}\right) + (\mathbf{y}-\mathbf{y}_{ij})^t \left(\frac{\partial \ln p(\tilde{\mathbf{x}}_{ij},\tilde{\mathbf{y}}_{ij})}{\partial \mathbf{y}}\right) \\ &\leq \left(\|\mathbf{x}-\mathbf{x}_{ij}\|^2 + \|\mathbf{y}-\mathbf{y}_{ij}\|^2\right)^{1/2} \left\|\frac{\partial \ln p(\tilde{\mathbf{x}}_{ij},\tilde{\mathbf{y}}_{ij})}{\partial(\mathbf{x},\mathbf{y})}\right\|\end{aligned}$$

for some vector $(\tilde{\mathbf{x}}_{ij},\tilde{\mathbf{y}}_{ij})$ on the line joining $(\mathbf{x},\mathbf{y})$ and $(\mathbf{x}_{ij},\mathbf{y}_{ij})$. Since $(\mathbf{x}_{ij},\mathbf{y}_{ij})$ is in the convex hull of $\mathcal{A}_N \cap T_i^x \times T_j^y$ we have $\|\mathbf{x}-\mathbf{x}_{ij}\| < 2r_x(\mathcal{A}_N)$ and $\|\mathbf{y}-\mathbf{y}_{ij}\| < 2r_y(\mathcal{A}_N)$. Thus,

$$D\left(p\|\tilde{p}^N\right) \leq 2(r_x(\mathcal{A}_N)^2 + r_y(\mathcal{A}_N)^2)^{1/2} \left(\sup_{\mathcal{A}_N} \|\frac{\partial \ln p(\mathbf{x},\mathbf{y})}{\partial(\mathbf{x},\mathbf{y})}\|\right)$$

The statement of the proposition follows from the fact that for any two distributions $d_V(P_1,P_2)^2 \leq 2D(P_1\|P_2)$ where $d_V(\cdot)$ denotes the variational distance [3]. □

Proposition 3 (Regarding (14a)) *Define the partial second moments outside* $\mathcal{A}_N$

$$\begin{aligned}m_x(\mathcal{A}_N^c) &\triangleq \int_{\mathcal{A}_N^c} \|\mathbf{x}-\hat{\mathbf{x}}_N\|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y} \\ m_y(\mathcal{A}_N^c) &\triangleq \int_{\mathcal{A}_N^c} \|\mathbf{y}-\hat{\mathbf{y}}_N\|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y}.\end{aligned}$$

Then

$$\left|E_{\tilde{p}^N}[\rho_N(\mathbf{X},\mathbf{Y})]\right| \leq (f_x(\mathcal{A}_N))^{1/2} (f_y(\mathcal{A}_N))^{1/2} + \left(\frac{m_x(\mathcal{A}_N^c)}{D_{x,N}}\right)^{1/2} \left(\frac{m_y(\mathcal{A}_N^c)}{D_{y,N}}\right)^{1/2} \tag{16}$$

where

$$\begin{aligned}f_x(\mathcal{A}_N) &\triangleq 4\frac{r_x(\mathcal{A}_N)^2}{D_{x,N}} \int_{\mathcal{A}_N} |\tilde{p}^N(\mathbf{x},\mathbf{y}) - p(\mathbf{x},\mathbf{y})| d\mathbf{x}d\mathbf{y} + 4\frac{r_x(\mathcal{A}_N)^2 P(\mathcal{A}_N^c)}{D_{x,N}} + \frac{m_x(\mathcal{A}_N^c)}{D_{x,N}} \\ f_y(\mathcal{A}_N) &\triangleq 4\frac{r_y(\mathcal{A}_N)^2}{D_{y,N}} \int_{\mathcal{A}_N} |\tilde{p}^N(\mathbf{x},\mathbf{y}) - p(\mathbf{x},\mathbf{y})| d\mathbf{x}d\mathbf{y} + 4\frac{r_y(\mathcal{A}_N)^2 P(\mathcal{A}_N^c)}{D_{y,N}} + \frac{m_y(\mathcal{A}_N^c)}{D_{y,N}}.\end{aligned}$$

Proof: The proof involves a tedious analysis starting from the definition of $E_{\tilde{p}^N}[\rho_N(\mathbf{X},\mathbf{Y})]$ and is relegated to the appendix.

Proposition 4 (Regarding (16))

1. $\frac{m_x(\mathcal{A}_N^c)}{D_{x,N}} \to 0$ *as* $P(\mathcal{A}_N^c) \to 0$.
2. $\frac{m_y(\mathcal{A}_N^c)}{D_{y,N}} \to 0$ *as* $P(\mathcal{A}_N^c) \to 0$.

Proof: Let

$$\begin{aligned} e_N^x(\mathbf{x},\mathbf{y}) &= \frac{\|\mathbf{x}-\hat{\mathbf{x}}_N\|}{\sqrt{D_{x,N}}}, \\ e_N^y(\mathbf{x},\mathbf{y}) &= \frac{\|\mathbf{y}-\hat{\mathbf{y}}_N\|}{\sqrt{D_{y,N}}}. \end{aligned}$$

Now

$$\begin{aligned} \frac{m_x(\mathcal{A}_N^c)}{D_{x,N}} &= \int_{\mathcal{A}_N^c} |e_N^x(\mathbf{x},\mathbf{y})|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}\, d\mathbf{y} \\ &\leq M^2 P(\mathcal{A}_N^c) + \sup_N \int_{\{e_N^x(\mathbf{x},\mathbf{y})>M\}} |e_N^x(\mathbf{x},\mathbf{y})|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}\, d\mathbf{y}\,. \end{aligned}$$

Similarly

$$\frac{m_y(\mathcal{A}_N^c)}{D_{y,N}} \leq M^2 P(\mathcal{A}_N^c) + \sup_N \int_{\{e_N^y(\mathbf{x},\mathbf{y})>M\}} |e_N^y(\mathbf{x},\mathbf{y})|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}\, d\mathbf{y}\,.$$

Hence to prove the proposition, it is sufficient to show that $\rho_N(\mathbf{x},\mathbf{y}), |e_N^x(\mathbf{x},\mathbf{y})|^2, |e_N^y(\mathbf{x},\mathbf{y})|^2$ are *uniformly integrable* for $N = 1, 2, \ldots$ [7]. By the Cauchy Schwarz inequality

$$\begin{aligned} E[|\rho_N(\mathbf{X},\mathbf{Y})|^{1+\delta}] &\leq \left(\frac{E[\|\mathbf{X}-\hat{\mathbf{X}}_N\|^{2+2\delta}]}{D_X^{1+\delta}}\right)^{1/2} \left(\frac{E[\|Y-\hat{\mathbf{Y}}_N\|^{2+2\delta}]}{D_Y^{1+\delta}}\right)^{1/2} \\ &\leq K' \end{aligned}$$

for some constant K' by the moment condition. Also by the same moment condition

$$\begin{aligned} E[|e_N^x(\mathbf{x},\mathbf{y})|^{2(1+\delta)}] &\leq K_1 \\ E[|e_N^y(\mathbf{x},\mathbf{y})|^{2(1+\delta)}] &\leq K_2 \end{aligned}$$

for some constants K_1, K_2. The required uniform integrability conditions follow by [7, p. 224, ex. 5.1]. □

Now we are ready to specify the sequences $\mathcal{A}_N$ and M_N. We pick

$$M_N = O(D_{x,N}^{-\epsilon} D_{y,N}^{-\epsilon}) \tag{17}$$

for some $\epsilon > 0$ and $\mathcal{A}_N$ as given Proposition 1.

Proposition 5 (Regarding (14c) and (14d)) *For M_N of (17) and $\mathcal{A}_N$ of Proposition 1, (14d) and (14c) are zero for large enough N.*

Proof: For $(\mathbf{x}, \mathbf{y}) \in \mathcal{A}_N$

$$\begin{aligned}
\frac{|\rho_N(\mathbf{x}, \mathbf{y})|}{M_N} &\leq \frac{1}{M_N} \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\sqrt{D_{x,N}}} \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|}{\sqrt{D_{y,N}}} \\
&\leq \frac{1}{M_N} \frac{r_x(\mathcal{A}_N)}{\sqrt{D_{x,N}}} \frac{r_y(\mathcal{A}_N)}{\sqrt{D_{y,N}}} \\
&= O(D_{x,N}^{\epsilon/2} D_{y,N}^{\epsilon/2}) \\
&< 1
\end{aligned}$$

for N sufficiently large. Thus the integration sets in (14c) and (14d) are empty as $N \to \infty$. □

Proof of the Theorem: The theorem is proved by arguing that each of the four terms in (14) tends to zero as $D_{x,N}, D_{y,N} \to 0$. We start with the (14a): From, (i) the choice of $\mathcal{A}_N$ as in Proposition 1; (ii)our assumption in (10) that $O(D_{x,N}) = O(D_{y,N})$; (iii) the bound in Proposition (2); we have,

$$\begin{aligned}
\int |p(\mathbf{x}, \mathbf{y}) - \tilde{p}^N(\mathbf{x}, \mathbf{y})| d\mathbf{x} d\mathbf{y} &\leq \sqrt{O(D_{x,N}^{\frac{1-\epsilon}{2}})} O(D_{x,N}^{-\epsilon(1+\delta)}) \\
&= O(D_{x,N}^{\frac{1}{4} - \epsilon(\frac{5}{4}+\delta)}).
\end{aligned}$$

Combining this with the bound in Proposition 3 and the limit in Proposition 4 gives

$$\begin{aligned}
f_x(\mathcal{A}_N) &= \frac{O(D_{x,N}^{1-\epsilon})}{D_{x,N}} O(D_{x,N}^{\frac{1}{4} - \epsilon(\frac{5}{4}+\delta)}) + \frac{O(D_{x,N}^{1-\epsilon})}{D_{x,N}} O(D_{x,N}^{\epsilon(1+\delta)}) + o(D_{x,N}) \qquad (18) \\
&\to 0 \qquad (19)
\end{aligned}$$

for $0 < \epsilon < 1/(9/4 + \delta)$. Similarly it can be shown that $f_y(\mathcal{A}_N) \to 0$. Thus we have shown that (14a) vanishes by (i) choice of $\mathcal{A}_N$ as in Proposition 1; (ii) bound in Proposition 3; and (iii) limits in Proposition 4.

Next, consider (14b):

$$\begin{aligned}
M_N \int_{\mathcal{A}_N} \left| p(\mathbf{x}, \mathbf{y}) - \tilde{p}^N(\mathbf{x}, \mathbf{y}) \right| d\mathbf{x} d\mathbf{y} &\overset{(a)}{\leq} 2M_N (r_x(\mathcal{A}_N)^2 + r_y(\mathcal{A}_N)^2)^{1/4} \left(\sup_{\mathcal{A}_N} \| \frac{\partial \ln p(\mathbf{x}, \mathbf{y})}{\partial (\mathbf{x}, \mathbf{y})} \| \right)^{1/2} \\
&\overset{(b)}{\leq} O(D_{x,N}^{-\epsilon} D_{y,N}^{-\epsilon}) O((D_{x,N}^{1-\epsilon} + D_{y,N}^{1-\epsilon})^{1/4}) O((D_{x,N} D_{y,N})^{-\epsilon(1+\delta)}) \\
&\to 0
\end{aligned}$$

where (a) is by Proposition 2 and (b) is by the choice of the sequence M_N as in (17), sequence $\mathcal{A}_N$ as in Proposition 1, and the limit follows by choosing ϵ sufficiently small.

The third and the fourth terms (14c) and (14d) vanish by Proposition 5. Thus we have shown that all the four terms in (14) vanish for appropriate choice of $M_N, \mathcal{A}_N$ and sufficiently small ϵ. Thus the theorem is proved. □

Appendix

A. Conditions implying the moment condition

The following are stronger conditions that imply the moment condition.

1. *uniform boundedness*:
$$\sup_N \frac{\|\mathbf{x} - \hat{\mathbf{x}}_N\|}{\sqrt{D_{x,N}}} \leq M$$
and a similar condition for $\mathbf{Y}$ clearly imply the moment condition.

2. If $g(\mathbf{x})$ is an integrable function such that
$$\int g(\mathbf{x})^{1+\delta} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \infty$$
and
$$\left(\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\sqrt{D_{x,N}}} \right)^2 \leq g(\mathbf{x}),$$
then the moment condition for $\mathbf{X}$ follows.

B. Proof of Proposition 3

First separate the averaging over $\mathcal{A}_N$ and $\mathcal{A}_N^c$:

$$E_{\tilde{p}^N}[\rho_N(\mathbf{x},\mathbf{y})] = \frac{\int_{\mathcal{A}_N} (\mathbf{x} - \hat{\mathbf{x}}_{ij})^t (\mathbf{y} - \hat{\mathbf{y}}_{ij}) \tilde{p}^N(\mathbf{x},\mathbf{y}) d\mathbf{x} d\mathbf{y}}{\sqrt{D_{x,N} D_{y,N}}} + \frac{\int_{\mathcal{A}_N^c} (\mathbf{x} - \hat{\mathbf{x}}_{ij})^t (\mathbf{y} - \hat{\mathbf{y}}_{ij}) \tilde{p}^N(\mathbf{x},\mathbf{y}) d\mathbf{x} d\mathbf{y}}{\sqrt{D_{x,N} D_{y,N}}}. \tag{20}$$

Note that our partition into terms involving integration inside the region $\mathcal{A}_N$ and outside $\mathcal{A}_N$ is natural since the two terms vanish for different reasons. The first term would be identically zero if the reconstructions $\mathbf{x}_{ij}, \mathbf{y}_{ij}$ were the geometric centroids. However, they are not geometric centroids but are second moment centroids w.r.t to the density $p(\mathbf{x},\mathbf{y})$. Nevertheless, as the quantization becomes increasingly refined and the quantization cell size tends to zero the second moment centroid does converge to the geometric centroid. The second term vanishes essentially by our moment criterion which guarantees that there are no "bad" cells as the quantization becomes increasingly refined.

Now we consider each of the two terms separately and show that they are bounded by the two terms in the statement of the proposition.

$$\begin{aligned}\frac{\int_{\mathcal{A}_N}(\mathbf{x}-\hat{\mathbf{x}}_{ij})^t(\mathbf{y}-\hat{\mathbf{y}}_{ij})\tilde{p}^N(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{\sqrt{D_{x,N}D_{y,N}}} &= \\ \sum_{i,j}\frac{P(S_i^x\times S_j^y)}{\sqrt{D_{x,N}D_{y,N}}}\frac{\int_{S_i^x\times S_j^y}(\mathbf{x}-\hat{\mathbf{x}}_{ij})^t(\mathbf{y}-\hat{\mathbf{y}}_{ij})d\mathbf{x}d\mathbf{y}}{\mathcal{V}(S_i^x\times S_j^y)} &\\ = \sum_{i,j}\frac{P(S_i^x\times S_j^y)}{\sqrt{D_{x,N}D_{y,N}}}a_{ij}^t b_{ij} &\end{aligned}$$

where we have defined

$$\begin{aligned} S_i^x\times S_j^y &\triangleq \mathcal{A}_N\cap T_i^x\times T_j^y \\ a_{ij} &\triangleq \frac{\int_{S_i^x}(\mathbf{x}-\hat{\mathbf{x}}_{ij})d\mathbf{x}}{\mathcal{V}(S_i^x)} \\ b_{ij} &\triangleq \frac{\int_{S_j^y}(\mathbf{y}-\hat{\mathbf{y}}_{ij})d\mathbf{x}}{\mathcal{V}(S_j^y)} \end{aligned}$$

and where $\mathcal{A}_N\cap T_i^x\times T_j^y$ is guaranteed to be a rectangle by the choice of $\mathcal{A}_N$ as in Proposition 1. Now we compute a_{ij}:

$$\begin{aligned} a_{ij} &= \frac{\int_{S_i^x}(\mathbf{x}-\hat{\mathbf{x}}_{ij})d\mathbf{x}}{\mathcal{V}(S_i^x)} \\ &= \frac{1}{\mathcal{V}(S_i^x)}\left[\int_{S_i^x}\mathbf{x}d\mathbf{x}-\hat{\mathbf{x}}_{ij}\mathcal{V}(S_i^x)\right] \\ &= \frac{\int_{S_i^x}\mathbf{x}d\mathbf{x}}{\mathcal{V}(S_i^x)}-\frac{\int_{T_i^x\times T_j^y}\mathbf{x}p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{P(T_i^x\times T_j^y)} \\ &= \frac{\int_{S_i^x\times S_j^y}\mathbf{x}\tilde{p}^N(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{P(S_i^x\times S_j^y)}-\frac{\int_{T_i^x\times T_j^y}\mathbf{x}p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{P(T_i^x\times T_j^y)} \\ &= \frac{1}{P(S_i^x\times S_j^y)}\left[\int_{S_i^x\times S_j^y}x[\tilde{p}^N(\mathbf{x},\mathbf{y})-\alpha_{ij}p(\mathbf{x},\mathbf{y})]d\mathbf{x}d\mathbf{y}-\alpha_{ij}\int_{\mathcal{A}_N^c\cap T_i^x\times T_j^y}xp(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}\right] \\ &= \frac{1}{P(S_i^x\times S_j^y)}\left[\int_{S_i^x\times S_j^y}(\mathbf{x}-\hat{\mathbf{x}}_{ij})[\tilde{p}^N(\mathbf{x},\mathbf{y})-\alpha_{ij}p(\mathbf{x},\mathbf{y})]d\mathbf{x}d\mathbf{y}\right] \\ &\quad -\frac{1}{P(S_i^x\times S_j^y)}\left[\alpha_{ij}\int_{\mathcal{A}_N^c\cap T_i^x\times T_j^y}(\mathbf{x}-\hat{\mathbf{x}}_{ij})p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}\right] \\ &= \frac{1}{P(S_i^x\times S_j^y)}\left[\Delta_x(i,j)+\delta_x(i,j)\right] \end{aligned}$$

where we have set

$$\begin{aligned} \alpha_{ij} &\triangleq \frac{P(S_i^x\times S_j^y)}{P(T_i^x\times T_j^y)} \\ \Delta_x(i,j) &\triangleq \int_{S_i^x\times S_j^y}(\mathbf{x}-\hat{\mathbf{x}}_{ij})[\tilde{p}^N(\mathbf{x},\mathbf{y})-\alpha_{ij}p(\mathbf{x},\mathbf{y})]d\mathbf{x}d\mathbf{y} \end{aligned}$$

$$\delta_x(i,j) \triangleq \alpha_{ij} \int_{\mathcal{A}_N^c \cap T_i^x \times T_j^y} (\mathbf{x} - \hat{\mathbf{x}}_{ij}) p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y} \ .$$

Similarly letting

$$\begin{aligned}
\Delta_y(i,j) &= \int_{S_i^x \times S_j^y} (\mathbf{y} - \hat{\mathbf{y}}_{ij})[\tilde{p}^N(\mathbf{x},\mathbf{y}) - \alpha_{ij} p(\mathbf{x},\mathbf{y})] d\mathbf{x}d\mathbf{y} \\
\delta_y(i,j) &= \alpha_{ij} \int_{\mathcal{A}_N^c \cap T_i^x \times T_j^y} (\mathbf{y} - \hat{\mathbf{y}}_{ij}) p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y}
\end{aligned}$$

we get

$$b_{ij} = \frac{1}{P(S_i^x \times S_j^y)} \left[\Delta_y(i,j) + \delta_y(i,j)\right]$$

To simplify notation set $P(S_i^x \times S_j^y) = \tilde{p}_{ij}$. Then by Cauchy Schwarz inequality

$$\begin{aligned}
&\sum_{i,j} \tilde{p}_{ij} \frac{a_{ij}^t}{\sqrt{D_{x,N}}} \frac{b_{ij}}{\sqrt{D_{y,N}}} \leq \\
&\qquad \left(\sum_{i,j} \tilde{p}_{ij} \| \frac{a_{ij}}{\sqrt{D_{x,N}}} \|^2 \right)^{1/2} \left(\sum_{i,j} \tilde{p}_{ij} \| \frac{b_{ij}}{\sqrt{D_{y,N}}} \|^2 \right)^{1/2} \\
&\leq \left(\sum_{i,j} \tilde{p}_{ij} \left(\frac{\|\Delta_x(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{x,N}}} + \frac{\|\delta_x(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{x,N}}} \right)^2 \right)^{1/2} \left(\sum_{i,j} \tilde{p}_{ij} \left(\frac{\|\Delta_y(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{y,N}}} + \frac{\|\delta_y(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{y,N}}} \right)^2 \right)^{1/2} \\
&\leq \left(2\sum_{i,j} \tilde{p}_{ij} \left(\frac{\|\Delta_x(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{x,N}}} \right)^2 + 2\sum_{i,j} \tilde{p}_{ij} \left(\frac{\|\delta_x(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{x,N}}} \right)^2 \right)^{1/2} \times \\
&\qquad \left(2\sum_{i,j} \tilde{p}_{ij} \left(\frac{\|\Delta_y(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{y,N}}} \right)^2 + 2\sum_{i,j} \tilde{p}_{ij} \left(\frac{\|\delta_y(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{y,N}}} \right)^2 \right)^{1/2} \qquad (21)
\end{aligned}$$

Now from the definition of $\delta_x(i,j)$,

$$\begin{aligned}
\frac{\|\delta_x(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{x,N}}} &\leq \alpha_{ij} \frac{\int_{\mathcal{A}_N^c \cap T_i^x \times T_j^y} \|x - \hat{\mathbf{x}}_{ij}\| p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y}}{\sqrt{D_{x,N}} P(S_i^x \times S_j^y)} \\
&\leq \frac{\alpha_{ij}}{P(S_i^x \times S_j^y)} \left(\int_{\mathcal{A}_N^c \cap T_i^x \times T_j^y} \|\mathbf{x} - \hat{\mathbf{x}}_{ij}\|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y} \right)^{1/2} \frac{\left(P(\mathcal{A}_N^c \cap T_i^x \times T_j^y) \right)^{1/2}}{\sqrt{D_{x,N}}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{i,j} \tilde{p}_{ij} \left(\frac{\|\delta_x(i,j)\|}{\tilde{p}_{ij}\sqrt{D_{x,N}}} \right)^2 &\leq \sum_{i,j} \alpha_{ij}^2 \left(\frac{\int_{\mathcal{A}_N^c \cap T_i^x \times T_j^y} \|\mathbf{x} - \hat{\mathbf{x}}_{ij}\|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y}}{D_{x,N}} \right) \frac{(1-\alpha_{ij}) P(T_i^x \times T_j^y)}{P(S_i^x \times S_j^y)} \\
&\leq \frac{\int_{\mathcal{A}_N^c} \|\mathbf{x} - \hat{\mathbf{x}}_{ij}\|^2 p(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y}}{D_{x,N}} = \frac{m_x(\mathcal{A}_N^c)}{D_{x,N}}. \qquad (22)
\end{aligned}$$

Now from the definition of $\Delta_x(i,j)$ we have,

$$\begin{aligned}
\|\Delta_x(i,j)\| &\leq \int_{S_i^x \times S_j^y} \|\mathbf{x} - \hat{\mathbf{x}}_{ij}\| \ |\tilde{p}^N(\mathbf{x},\mathbf{y}) - \alpha_{ij} p(\mathbf{x},\mathbf{y})| d\mathbf{x}d\mathbf{y} \\
&\leq r_x(\mathcal{A}_N) \int_{S_i^x \times S_j^y} |\tilde{p}^N(\mathbf{x},\mathbf{y}) - \alpha_{ij} p(\mathbf{x},\mathbf{y})| d\mathbf{x}d\mathbf{y}
\end{aligned}$$

where the last inequality is because of the definition of $r_x(\mathcal{A}_N)$. Hence

$$\begin{aligned}
\sum_{i,j}\tilde{p}_{ij}\left(\frac{\|\Delta_x(i,j)\|}{\sqrt{D_{x,N}}\tilde{p}_{ij}}\right)^2 &= 16\frac{r_x(\mathcal{A}_N)^2}{D_{x,N}}\sum_{i,j}\tilde{p}_{ij}\left(\frac{\int_{S_i^x\times S_j^y}|\tilde{p}^N(\mathbf{x},\mathbf{y})-\alpha_{ij}p(\mathbf{x},\mathbf{y})|d\mathbf{x}d\mathbf{y}}{4P(S_i^x\times S_j^y)}\right)^2 \\
&\overset{(a)}{\leq} 16\frac{r_x(\mathcal{A}_N)^2}{D_{x,N}}\sum_{i,j}P(S_i^x\times S_j^y)\left(\frac{\int_{S_i^x\times S_j^y}|\tilde{p}^N(\mathbf{x},\mathbf{y})-\alpha_{ij}p(\mathbf{x},\mathbf{y})|d\mathbf{x}d\mathbf{y}}{4P(S_i^x\times S_j^y)}\right) \\
&\leq 4\frac{r_x(\mathcal{A}_N)^2}{D_{x,N}}\sum_{i,j}\int_{S_i^x\times S_j^y}|\tilde{p}^N(\mathbf{x},\mathbf{y})-\alpha_{ij}p(\mathbf{x},\mathbf{y})|d\mathbf{x}d\mathbf{y} \\
&\leq 4\frac{r_x(\mathcal{A}_N)^2}{D_{x,N}}\left[\int_{\mathcal{A}_N}|\tilde{p}^N(\mathbf{x},\mathbf{y})-p(\mathbf{x},\mathbf{y})|d\mathbf{x}d\mathbf{y}+\sum_{i,j}(1-\alpha_{ij})P(S_i^x\times S_j^y)\right] \\
&\leq 4\frac{r_x(\mathcal{A}_N)^2}{D_{x,N}}\int_{\mathcal{A}_N}|\tilde{p}^N(\mathbf{x},\mathbf{y})-p(\mathbf{x},\mathbf{y})|d\mathbf{x}d\mathbf{y}+4\frac{r_x(\mathcal{A}_N)^2P(\mathcal{A}_N^c)}{D_{x,N}}. \qquad (23)
\end{aligned}$$

where (a) is because of $\sum_i p_i s_i^2 \leq \sum_i p_i s_i$ whenever $s_i \leq 1$.

Similarly the terms corresponding to y can also be bounded. Combining equations (22), (23), (21) and using the definitions of $f_x(\mathcal{A}_N)$ and $f_y(\mathcal{A}_N)$ we have

$$\frac{\int_{\mathcal{A}_N}(\mathbf{x}-\hat{\mathbf{x}}_{ij})^t(\mathbf{y}-\hat{\mathbf{y}}_{ij})\tilde{p}^N(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{\sqrt{D_{x,N}D_{y,N}}} \leq (f_x(\mathcal{A}_N))^{1/2}(f_y(\mathcal{A}_N))^{1/2} \qquad (24)$$

Now we consider the second term in equation (20)

$$\begin{aligned}
&\frac{\int_{\mathcal{A}_N^c}(\mathbf{x}-\hat{\mathbf{x}}_{ij})^t(\mathbf{y}-\hat{\mathbf{y}}_{ij})\tilde{p}^N(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{\sqrt{D_{x,N}D_{y,N}}} = \\
&= \frac{\int_{\mathcal{A}_N^c}(\mathbf{x}-\hat{\mathbf{x}}_{ij})^t(\mathbf{y}-\hat{\mathbf{y}}_{ij})p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{\sqrt{D_{x,N}D_{y,N}}} \\
&\overset{(b)}{\leq} \left(\frac{\int_{\mathcal{A}_N^c}\|\mathbf{x}-\hat{\mathbf{x}}_{ij}\|^2p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{D_{x,N}}\right)^{1/2}\left(\frac{\int_{\mathcal{A}_N^c}\|\mathbf{y}-\hat{\mathbf{y}}_{ij}\|^2p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}}{D_{y,N}}\right)^{1/2} \\
&= \left(\frac{m_x(\mathcal{A}_N^c)}{D_{x,N}}\right)^{1/2}\left(\frac{m_y(\mathcal{A}_N^c)}{D_{y,N}}\right)^{1/2}
\end{aligned}$$

where (b) is by Cauchy Schwarz inequality. The proposition follows from the last inequality and equations (24) and (20). □

Acknowledgements

We thank Toby Berger for discussions that motivated this work and thank Tamas Linder for his comments on an early version of this paper that helped greatly improve its quality.

References

[1] T. Berger. *Multiterminal Source Coding.* In G.Longo, editor, the Information Theory Approach to Communications, Springer-Verlag, New York, 1977.

[2] R. E. Blahut. *Principles and Practice of Information Theory.* Addison Wesley, Reading, MA, 1987.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[4] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Trans. Information Theory*, IT-37:1501–1518, Nov. 1991.

[5] A. Gersho. Asymptotically optimal block quantization. *IEEE Trans. Information Theory*, IT-25:373–380, July 1979.

[6] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic Pub., Boston, 1992.

[7] R. Durrett. *Probability- Theory and Examples.* Wadsworth Brooks/Cole Pub., California, 1991.

[8] N. S. Jayant and P. Noll. *Digital Coding of Waveform.* Prentice-Hall, Englewood Cliffs, NJ, 1984.

[9] T. Linder and R. Zamir. On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Information Theory*, IT-40:2026–2031, Nov. 1994.

[10] T. Linder, R. Zamir, and K. Zeger. Multidimensional companding for non-difference distortion measures:the rate distortion function *IEEE Trans. Information Theory*, IT-45:533-547, March 1999.

[11] T. Linder, R. Zamir, and K. Zeger. The multiple description rate region for high resolution source coding. In *Proc. of Data Compression Conf.*, pages 149–158, Snowbird, Utah, March 1998.

[12] T. Linder and K. Zeger. Asymptotic entropy constrained performance of tessellating and universal randomized lattice quantization. *IEEE Trans. Information Theory*, pp. 575-579, March 1994.

[13] S. Na and D.L. Neuhoff. Bennett's integral for vector quantizers. *IEEE Trans. Information Theory*, IT-41:886–900, July. 1995.

[14] W. Rudin. *Principles of Mathematical Analysis*, Page 111. McGraw Hill Pub., New York.

[15] W. Rudin. *Real and Complex Analysis*, . McGraw Hill Pub., New York.

[16] H. Viswanathan and T. Berger. The quadratic Gaussian CEO problem. *IEEE Trans. Information Theory*, IT-43:1549–1559, Sept. 1997.

[17] R. Zamir and T. Berger. Multiterminal source coding with high resolution. *IEEE Trans. Information Theory*, IT-45:106–117.

[18] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Information Theory*, IT-42:1152-1159, July 1996.